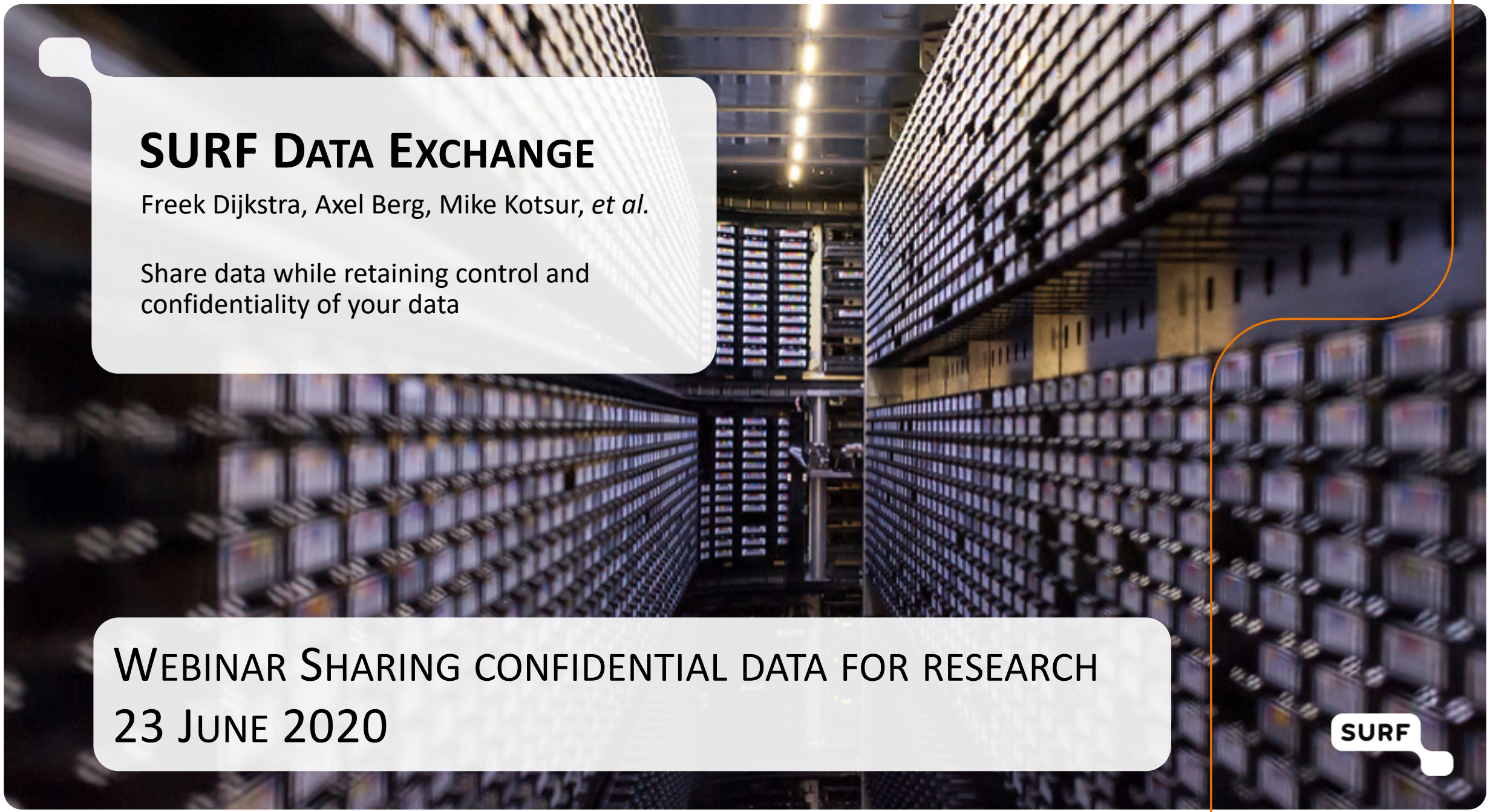


SURF Data Exchange

Freek Dijkstra, Axel Berg, Mike Kotsur, *et al.*

Share data while retaining control and confidentiality of your data

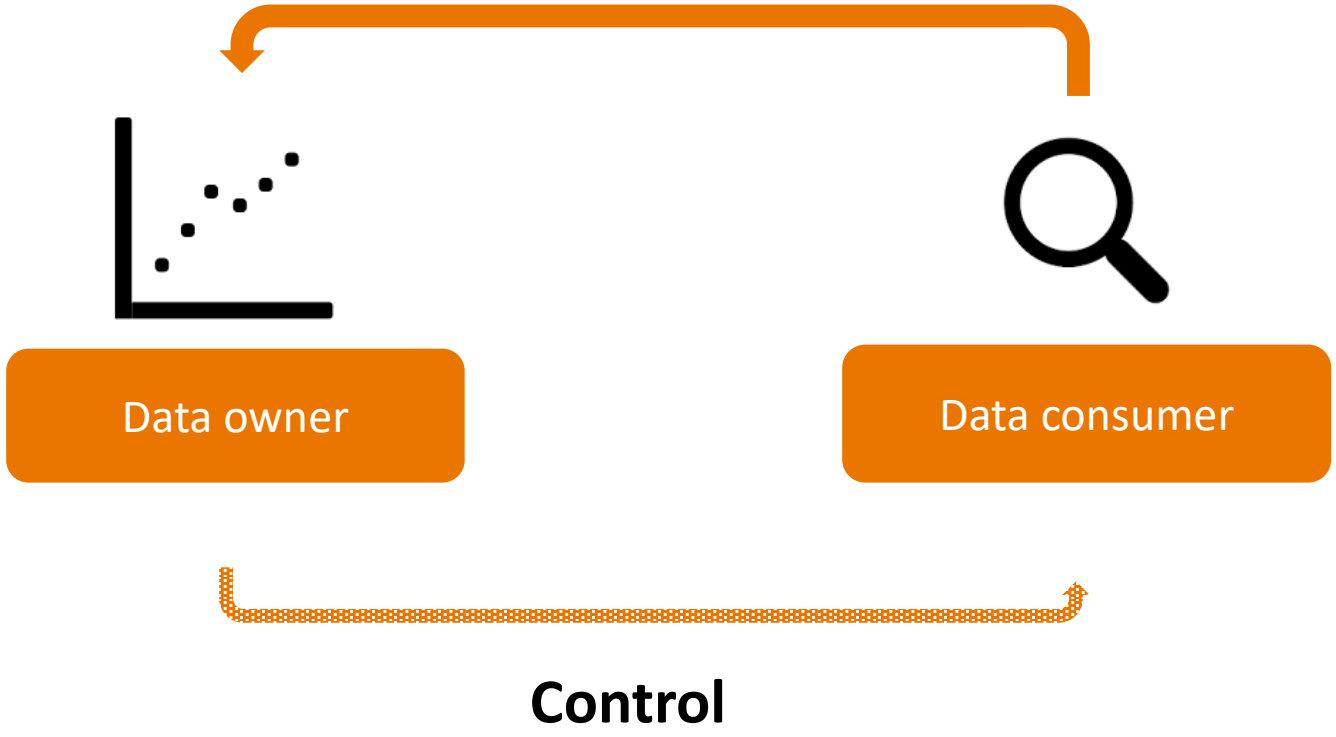
Webinar Sharing confidential data for research
23 June 2020

SURF

Version 2020-06-22

Barrier for data sharing

⚠ Gain is usually with the data consumer, burden is with the data provider



Willingness to share data

ROI

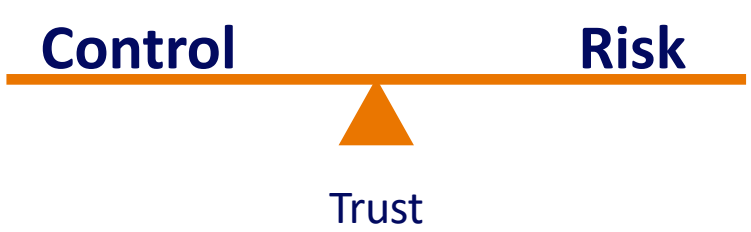
Return on Investment (ROI) is determined by the balance between effort it takes to share data, and the gain received by sharing data

Return on investment

+

Trust

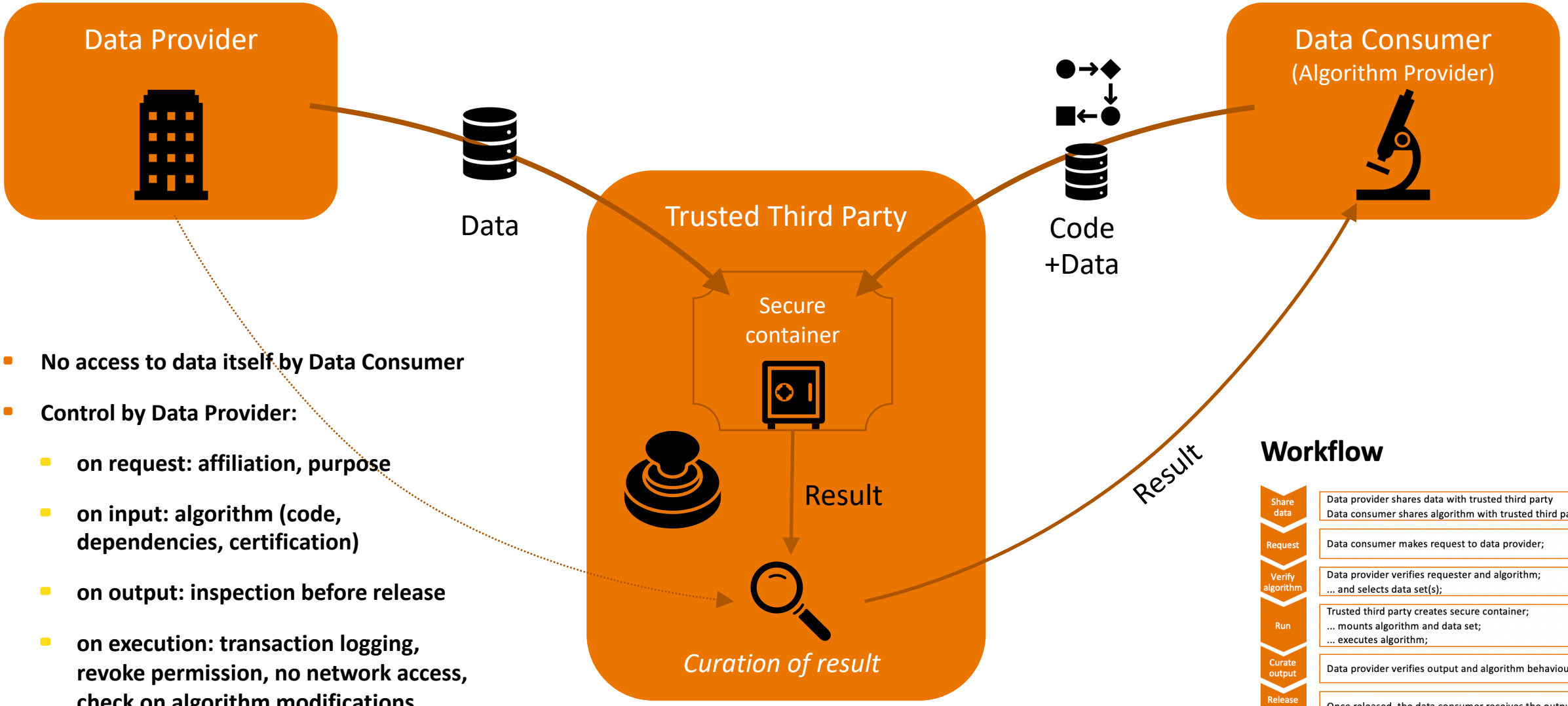
Trust is determined by the balance between the risks (due to privacy or competition), and the control (due to verification and security) of sharing and usage of data



Trust

Original diagram by Nadia Piet, Ocean Conijn, and Joris van Rossum

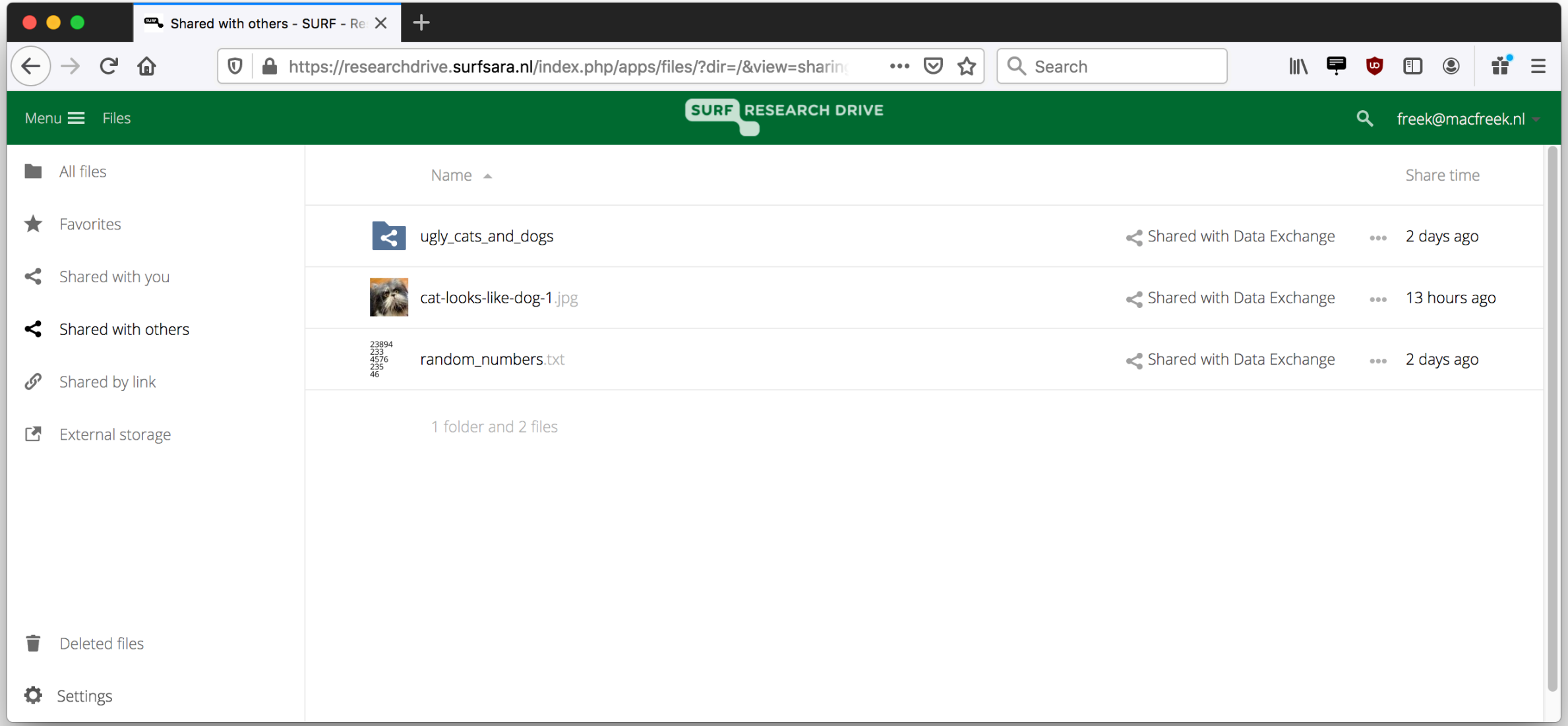
Working prototype on trusted data sharing



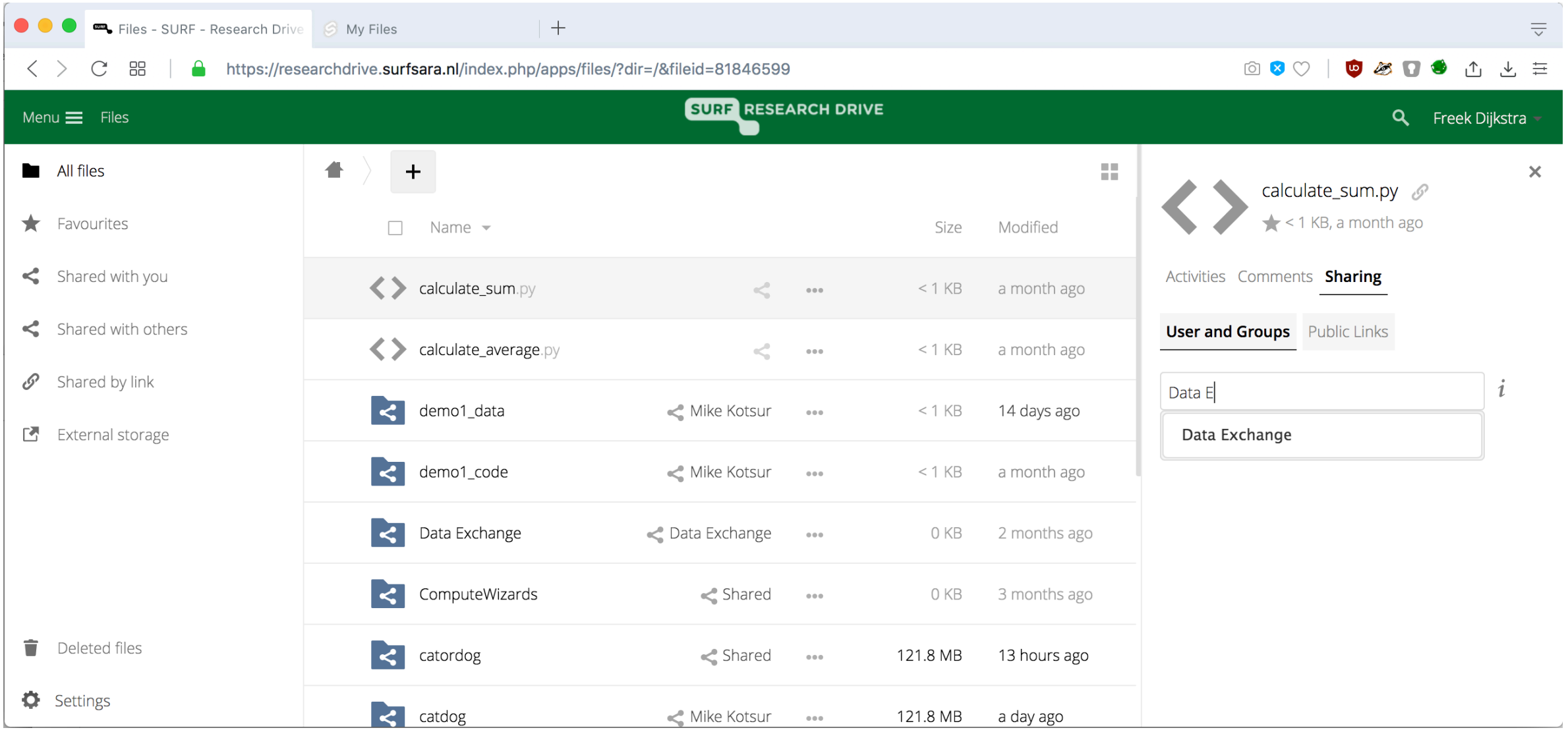
- **No access to data itself by Data Consumer**
- **Control by Data Provider:**
  - **on request: affiliation, purpose**
  - **on input: algorithm (code, dependencies, certification)**
  - **on output: inspection before release**
  - **on execution: transaction logging, revoke permission, no network access, check on algorithm modifications**

Workflow

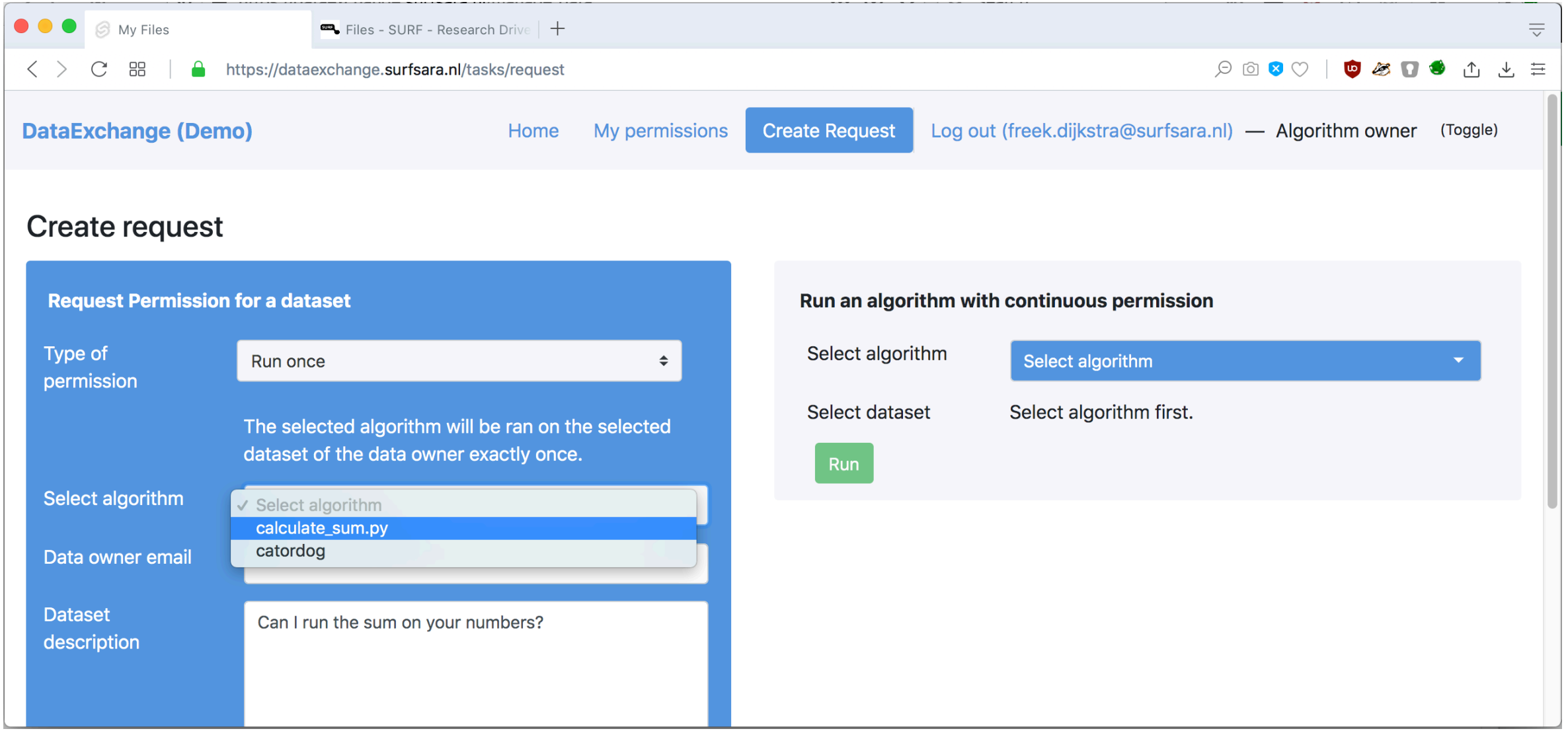
Share data	Data provider shares data with trusted third party Data consumer shares algorithm with trusted third party
Request	Data consumer makes request to data provider;
Verify algorithm	Data provider verifies requester and algorithm; ... and selects data set(s);
Run	Trusted third party creates secure container; ... mounts algorithm and data set; ... executes algorithm;
Curate output	Data provider verifies output and algorithm behaviour;
Release output	Once released, the data consumer receives the output.



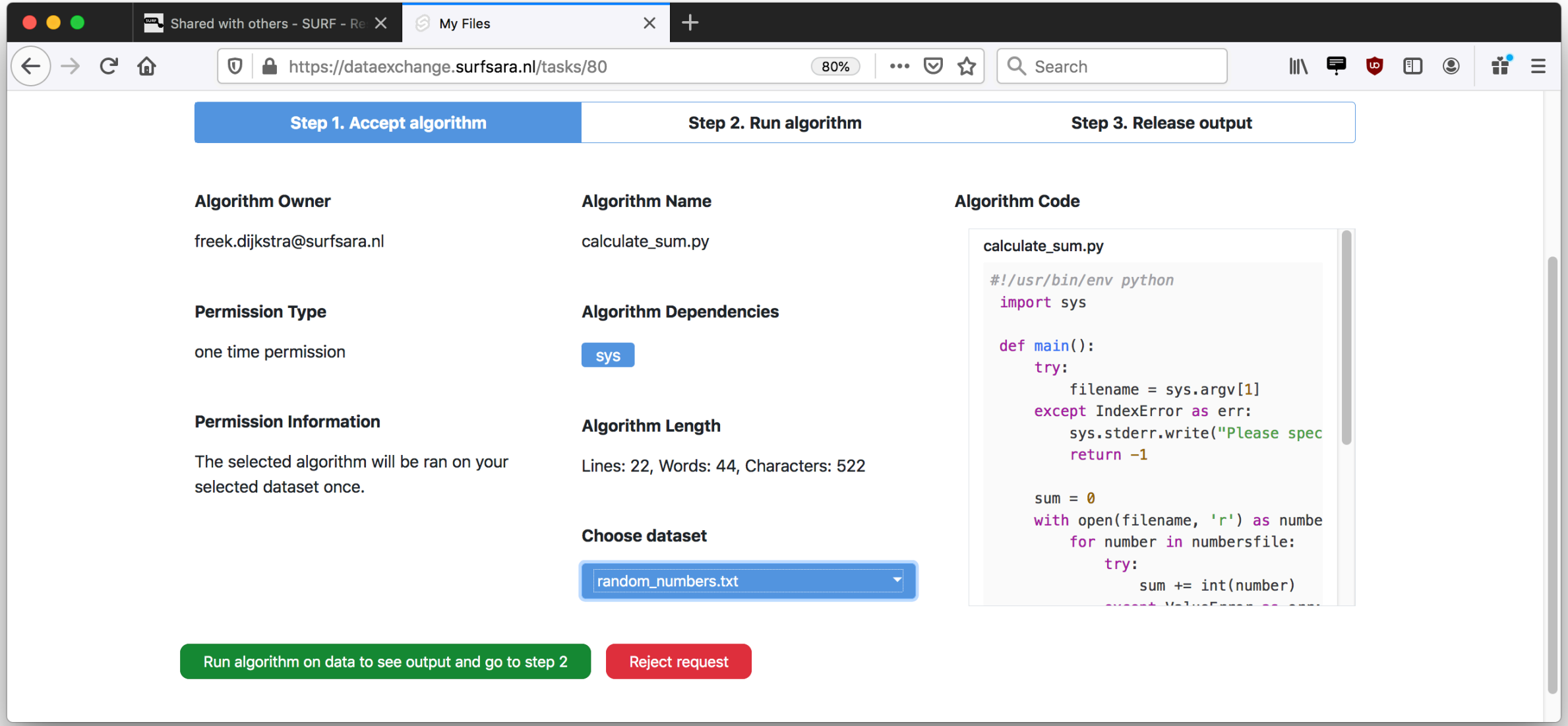
Data is shared with the Data exchange



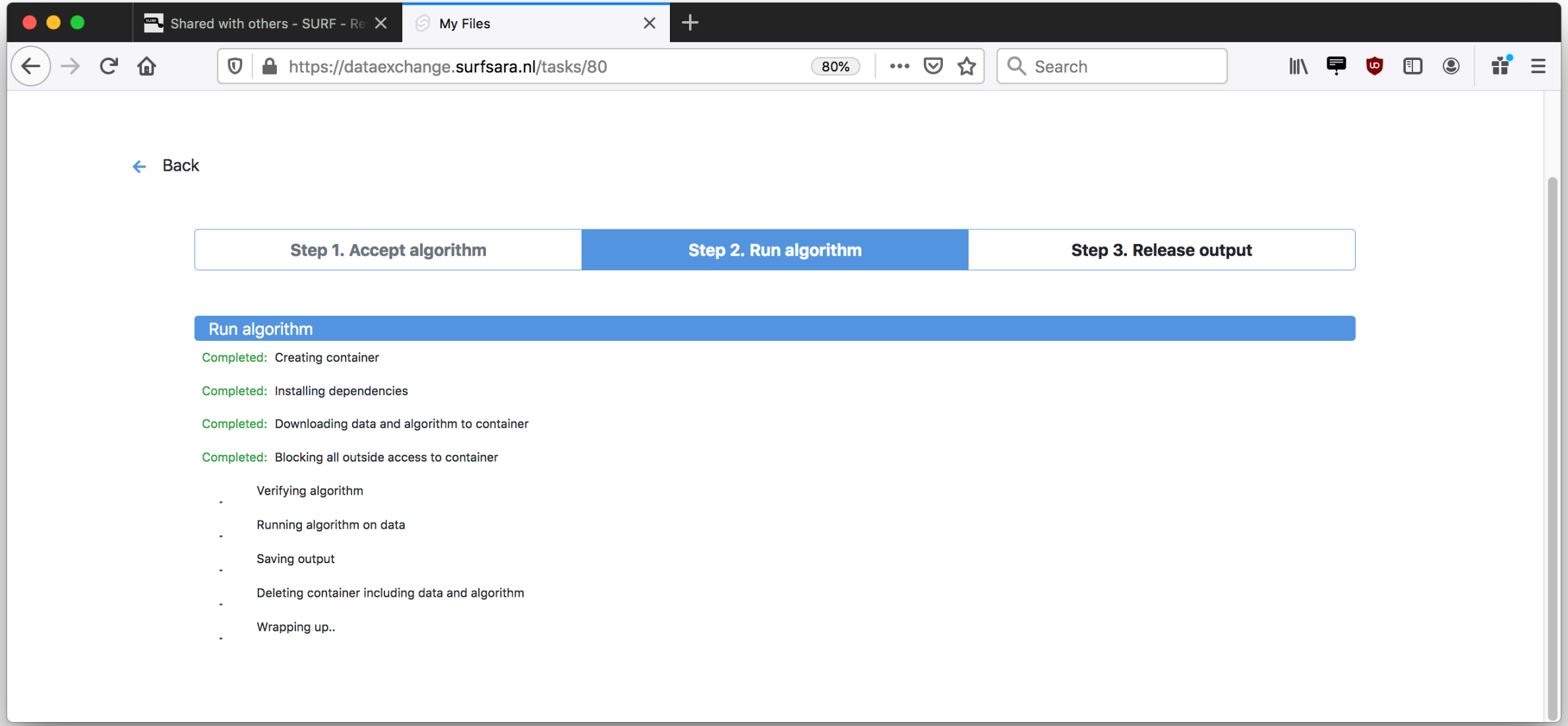
Algorithm is shared with the Data exchange by researcher



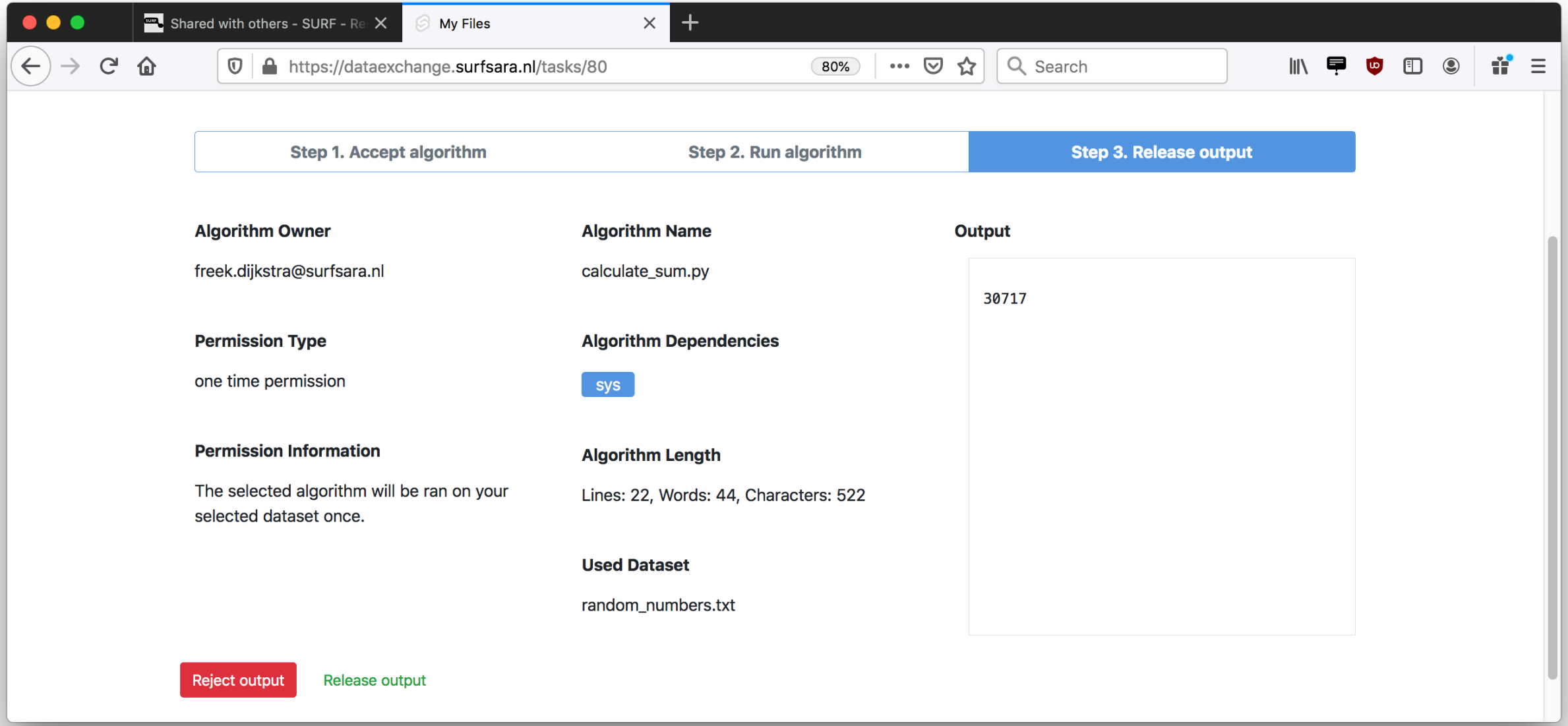
Researcher makes a request to the data provider



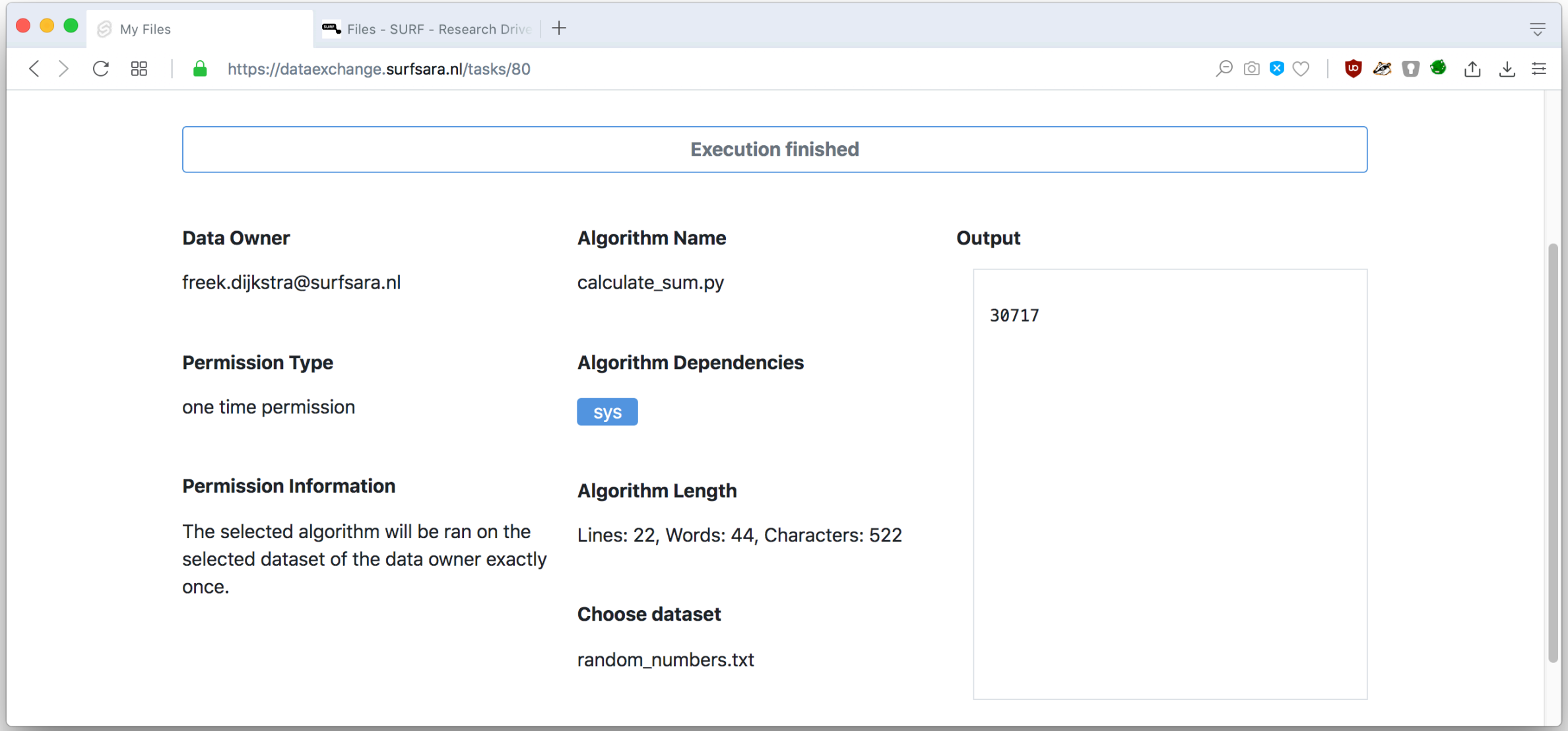
Data provider reviews request and selects dataset



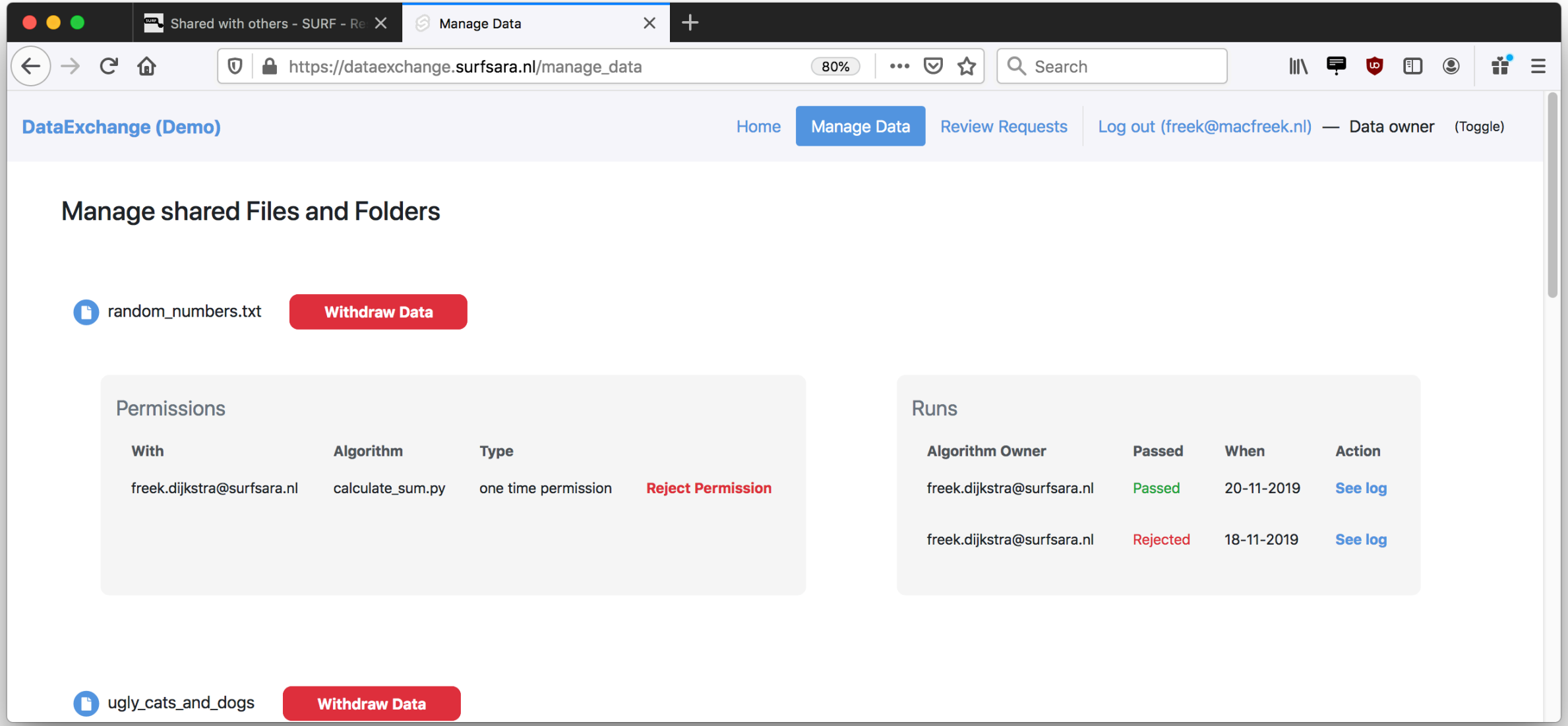
Trusted Third Party runs algorithm on dataset



Data provider reviews output



Researcher can see released output



Data provider can at any time withdraw permissions

Next Steps

- Better understanding of the needs and requirements
  - This webinar!
- Work with potential pilot partners
  - Integrate with ODISSEI Data Node
  - Talks with other interested organisations (perhaps you?)
- Make the demo accessible to everyone.
  - Currently requires a SURF ResearchDrive account
  - We'll make it work with e.g. Google Drive (August 2020)

Different Methods to Ease Data Sharing

Agreements

- Stipulation of what can/cannot be done
- Signing of contract or NDA
- Dispute resolution process

Registration

- Authentication
- Verification of credential
- Reputation score
- Policy framework
- Audit trails

Pseudonymization

- Filtering (on records)
- Pruning (on properties)
- Aggregation (combine records)
- Make coarse grained buckets
- Slight alteration of data
- One-way hashing
- One-time identifiers
- Synthetic data (mix records / AI)

Data Vault

- Data source retains control
- Delegate permissions
- No central data lake
- Data marketplace

Secure Containers

- Bring algorithm to data
- At Trusted third party or at data provider
- Share output instead of data

Secure Computing

- Secure multi-party computation
- Homomorphic encryption
- Garbled Circuits
- Zero-knowledge proof

COLLABORATION WITHOUT SHARING DATA

Freek Dijkstra

Freek.Dijkstra@surfsara.nl

www.surf.nl

This presentation is available under the creative commons attribution 4.0 license

Driving innovation together

SURF